
REJOINDER TO ISSUE 12(3)

Formative Versus Reflective Measurement of Executive Function Tasks: Response to Commentaries and Another Perspective

Michael T. Willoughby

RTI International

The focus article (Willoughby et al., 2014) (1) introduced the distinction between formative and reflective measurement and (2) proposed that performance-based executive function tasks may be better conceptualized from the perspective of formative rather than reflective measurement. This proposal stands in sharp contrast to conventional measurement wisdom, in which confirmatory factor models are routinely used to represent individual differences in executive function ability. Here, I respond to the many thoughtful commentaries on my proposal. In addition, I draw on a philosophical distinction between formative and reflective latent variables that was made by Borsboom, Mellenbergh, and van Heerden (2003) to further challenge the current tendency to view executive function tasks from the perspective of reflective measurement.

RESPONSE TO ENGELHARD AND WANG

Engelhard and Wang (2014) raised 3 ideas that merit some comment. First, the authors stated:

The purpose of formative models is to create a composite index to represent EF, and the purpose of reflective models is to create a latent variable to represent EF (p. 104).

Although this may be a widely held opinion, it contradicts ideas that were introduced by Bollen and Bauldry (2011), who made a distinction between causal and composite

This is a rejoinder to the commentaries on “Executive Functions: Formative Versus Reflective Measurement” that published in issue 12(3) 2014 of *Measurement: Interdisciplinary Research & Perspectives*.

Correspondence should be addressed to Michael T. Willoughby, RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709-2194. E-mail: mwilloughby@rti.org

indicators. Following their notation, an equation that represented causal indicators took the form

$$\eta_{1i} = \alpha_\eta + \gamma_{11}x_{1i} + \gamma_{12}x_{2i} + \dots + \gamma_{1Q}x_{Qi} + \zeta_{1i},$$

where η was a *latent* variable for the i th case, x_{qi} was the q th observed (causal) indicator, and ζ was a disturbance term that represented all of the other variables that cause the latent variable, η . Bollen and Bauldry (2011) posited that causal indicators should have “conceptual unity,” meaning the causal indicators should conform to the theoretical definition of a latent variable. They proposed an alternative equation for composite indicators that took the form

$$C_{1i} = w_{10} + w_{11}x_{1i} + w_{12}x_{2i} + \dots + w_{1Q}x_{Qi},$$

where C is a *composite* variable for the i th case and x_{qi} is the q th observed (composite) indicator. Notably, there is no disturbance term because composite variables are exact linear combinations of their indicators. Moreover, composite indicators are not assumed to have conceptual unity because the purpose of composite variables is simply to combine a set of scores into a summary variable. The important point is that formative models can be used to create either latent or composite variables (i.e., *latent* variables are not limited to reflective measurement).

Second, Engelhard and Wang elaborated the reasons why they prefer reflective over formative models (i.e., exchangeability and invariance of indicators, ability to establish nomological networks). Although I share many of their opinions, personal preferences are not a sufficient criterion for determining whether performance-based tasks are best construed as causal, composite, or effect indicators of the latent construct of EF. Although causal indicator models contradict conventional measurement wisdom and introduce a number of analytic challenges (e.g., identification problems, strategies for evaluating measurement invariance), these limitations must be weighed against the potential errors in inference that result from the application of standard analytic approaches (which are all predicated on reflective measurement) to data that may not conform to their underlying assumptions.

Third, Engelhard and Wang (2014) expressed enthusiasm for the use of item response theory (IRT) models to EF data. My colleagues and I share similar enthusiasm and have outlined some of the merits of using IRT models to evaluate individual EF tasks elsewhere (Willoughby, Wirth, & Blair, 2011). It is important to point out that the primary question posed in the focus article was how best to represent individuals’ performance across a battery of EF tasks. I do not believe that either Rasch or IRT models can assist with this problem. Although both Rasch and IRT models can substantially improve the evaluation of individual EF tasks, they continue to assume reflective measurement in that an underlying latent variable (θ) is assumed to give rise to observed item level data.

RESPONSE TO ROOS

Many substantive researchers and applied data analysts who work with EF data will likely not have familiarity with vanishing tetrad tests (VTTs). Roos (2014) provided both a nice synopsis

of VTTs, as well as a reasonable critique of my colleagues and my use of VTTs in the focus article. Whereas (completely) effect indicator models assume that all model-implied tetrads vanish, (completely) causal indicator models assume that none of the model-implied tetrads vanish. Hence, the statistical significance of the VTT chi-square test provides one means for contrasting these models. However, as Roos correctly noted, the VTT chi-square-test statistic is simply an indicator of model fit. Models that exhibited poor fit (i.e., significant VTT test statistics) may have fit poorly for reasons other than the distinction between (completely) causal and (completely) effect indicators that we emphasized. Roos further noted that VTTs may be most useful in formal model comparisons; that is, many chi-square equivalent models are nested with respect to their vanishing tetrads. Nested VTTs provide 1 means of contrasting competing models. I agree with Roos that this is an important direction for future research, and it is something that my colleagues and I are actively pursuing. Unfortunately, formal model comparisons of this sort were not feasible in the focus article because we relied entirely on a reanalysis of published data. Minimally, we hope that our focus article stimulated greater interest in the use of VTTs as an empirical approach for informing questions of formative versus reflective measurement problems.

RESPONSE TO EID AND KOCH

In the focus article, my colleagues and I considered plausibility questions (attributable to Jarvis, MacKenzie, & Podsakoff, 2003) and vanishing tetrad tests (attributable to Bollen and colleagues) as 2 approaches for determining whether performance-based tasks were best conceived of as causal or effect indicators of the latent construct of EF. Eid and Koch (2014) introduced a third approach—namely, a reliance on stochastic measurement theory—for evaluating the question of formative versus reflective measurement of EF. Given my lack of familiarity with formal stochastic measurement theory, I had a difficult time critically evaluating their contribution, though was relieved that they were in general agreement with the ideas raised in the focus article (though for entirely different reasons). To the extent that there is convergence in conclusions across different approaches, this provides greater confidence in any resulting conclusion. Eid and Koch also proposed the possibility of using a profile-oriented approach for future studies of EF. Here, the focus is on arraying individuals into homogenous subgroups. Although this approach will not be applicable for all questions, this represents an underutilized strategy in this area of study.

RESPONSE TO WIEBE & MCFALL

Unlike the other commentators, Wiebe has utilized CFA methods with EF data in previously published papers (Schoemaker et al., 2012; Wiebe, Espy, & Charak, 2008; Wiebe et al., 2011). As such, I was particularly interested in her commentary. I acknowledge that thinking about performance-based tasks as causal indicators of EF is somewhat foreign and potentially confusing. Although SES is the prototypical example used in the literature of formative measurement, Wiebe and McFall (2014) noted problems in this example from their perspective. Their nomination of *functional age* as an alternative example was welcome. To the extent that functional age is a construct that both derives its meaning from its indicators and takes on different meaning depending on the indicators chosen, it meets 2 of the defining characteristics of formative

models. Podsakoff, MacKenzie, Podsakoff, and Lee (2003) provided a number of examples of constructs in the area of leadership research that served as exemplars of formative constructs (e.g., charismatic leadership; transformational leadership; empathetic leadership).

Wiebe and McFall correctly noted that irrespective of whether EF is best construed as a formative or reflective construct (or defined by some combination of causal and effect indicators), practically speaking, both perspectives benefit from a greater number of indicators (even though the rationale for an increased number of indicators may differ). I completely agree with this insight and acknowledge that in some settings and/or for some populations of participants, obtaining multiple EF tasks is challenging. However, the essential concern is not with how many indicators are optimal to measure a latent variable but rather with how one goes about summarizing performance across the measures that one has.

Finally, similar to Engelhard and Wang, Wiebe and McFall raised concerns about the practical problems that arise if EF is a formative construct. In addition to raising issues regarding statistical identification and the apparent irrelevance of testing the dimensionality (factor structure) of EF that were raised by Wiebe and McFall, I would also note that some forms of reliability are no longer relevant from the perspective of formative measurement. For example, my colleagues and I recently demonstrated how changes in maximal reliability could be used to construct short forms of our EF task battery (Willoughby, Pek, & Blair, 2013). However, this work was entirely predicated on assumptions of reflective measurement. If EF is a formative construct, the rationale for this previous work no longer holds. In general, many well-established psychometric techniques (which collectively assume reflective measurement) that address important issues are irrelevant for formative constructs. This is unsettling and underscores the importance of multiple research groups investigating this issue.

ONTOLOGICAL STATUS OF LATENT VARIABLES —EVIDENCE FOR FORMATIVE MODELS OF EF?

Borsboom, Mellenbergh, and van Heerden (2003) provided a thought-provoking evaluation of the theoretical status of latent variables from the perspective of modern test theory. They organized a portion of their essay around 3 questions regarding the presumed relationship between latent variables and their observed indicators. In particular, they asked (1) whether the relationship between latent variables and their indicators was causal, (2) if so, in what sense, and (3) whether latent variable theory was neutral with respect to these issues (Borsboom et al., 2003; p. 204). Although space constraints limit a full consideration of their ideas, they concluded that the decision regarding whether a latent construct was formative or reflective was dependent on the presumed ontological status of latent variables. Reflective models imply a realist philosophical view in which latent variables are presumed to exist apart from and precede the measurement of indicator variables (e.g., people have some true level of inhibitory control, which gives rise to their performance on antisaccade, stoop, and stop signal tasks). In contrast, formative models imply a constructivist philosophical view in which latent variables do not exist apart from observed measures but instead reflect a summary of such measures (e.g., inhibitory control is defined by a person's performance on antisaccade, stoop, and stop-signal tasks). Hence, consideration of the ontological status of latent variables represents another approach for helping to

discern whether EF (or its subconstructs: inhibitory control, working memory, attention shifting) is best conceived of as a formative or reflective latent variable.

Historically, the prefrontal cortex was identified as the neural substrate that was responsible for cognitive functions that resemble modern EF (Luria, 1966, 1973). This, in turn, influenced seminal ideas in neuropsychological research, including the introduction of the central executive and the supervisory attention systems (Baddeley, 1986; Norman & Shallice, 1986; Shallice, 1982). Although the idea that EFs were localized to the prefrontal cortex served a valuable heuristic function, modern characterizations of EF (and its subcomponents) have replaced notions of localization (and regional specialization) with a more distributed perspective (e.g., Andres, 2003; Baddeley, 1998; Collette & Van der Linden, 2002; GoldmanRakic, 1996). For example, in the case of inhibitory control, Munakata et al. (2011) proposed that the prefrontal cortex was specialized for actively representing and maintaining abstract information and that these representations facilitated 2 different types of inhibitory control that influenced other parts of the brain. As they noted, “What distinguishes different prefrontal regions and their roles in distinct types of inhibition is the nature of their connectivity with other brain regions and the content of the abstract information represented” (p. 457). Similarly, in a review of cognitive and neural models of working memory, D’Esposito (2007) concluded that “working memory can be viewed as neither a unitary nor dedicated system. A network of brain regions, including the [prefrontal cortex], is critical for the active maintenance of internal representations that are necessary for goal-directed behavior” (p. 769). Similar conclusions resulted from computational models of working memory (Hazy, Frank, & O’Reilly, 2006). Consistent with these ideas, EF may be profitably understood to represent an *emergent property* of individuals that results from interactions between neural systems that support goal-directed behaviors (Buss & Spencer, 2014; D’Esposito, 2007; GoldmanRakic, 1996; Maniadakis, Trahanias, & Tani, 2012).

To the extent that EF is an emergent property that is defined by interactions between neural systems, this would appear to favor a constructivist (formative) over realistic (reflective) account of the latent variable EF. To be clear, although neither formative nor reflective latent variables may adequately capture the variety or complexity of cognitive processes that support goal-directed behaviors, the suggestion here is that formative latent variables may provide a better approximation to modern theoretical accounts of EF than do reflective latent variables. Similar to our consideration of Jarvis et al.’s (2003) plausibility questions in the focus article, appealing to differences in the presumed ontological status of latent variables is a necessarily subjective process for which no definitive conclusions can be drawn. Nonetheless, the resolution of whether EF tasks are best characterized as a formative or reflective latent variable will likely require a combination of theoretical considerations and empirical data. Ultimately, the resolution of whether EF is best construed as a formative versus reflective latent variable is only important to the extent that this distinction influences substantive conclusions that are drawn regarding the presumed developmental causes, course, and consequences of EF across the lifespan.

FUNDING

This research was supported by the Institute of Education Sciences, U.S. Department of Education (R324A120033). The opinions expressed are those of the author and do not represent views of this funding agency.

REFERENCES

Andres, P. (2003). Frontal cortex as the central executive of working memory: Time to revise our view. *Cortex*, 39(4–5), 871–895. doi:10.1016/S0010-9452(08)70868-2

Baddeley, A. (1998). The central executive: A concept and some misconceptions. *Journal of the International Neuropsychological Society*, 4(5), 523–526.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press.

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, 16(3), 265–284. doi:10.1037/A0024448

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203–219. doi:10.1037/0033-295x.110.2.203

Buss, A. T., & Spencer, J. P. (2014). The emergent executive: A dynamic field theory of the development of executive function. *Monographs of the Society for Research in Child Development*, 79(2), vii, 1–103. doi:10.1002/mono.12096

Collette, F., & Van der Linden, M. (2002). Brain imaging of the central executive component of working memory. *Neuroscience & Biobehavioral Reviews*, 26(2), 105–125.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 362(1481), 761–772. doi:10.1098/rstb.2007.2086

GoldmanRakic, P. S. (1996). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 351(1346), 1445–1453. doi:10.1098/rstb.1996.0129

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: Making working memory work. *Neuroscience*, 139(1), 105–118. doi:10.1016/j.neuroscience.2005.04.067

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, 30(2), 199–218. doi:10.1086/376806

Luria, A. R. (1966). *Higher cortical functions in man*. New York, NY: Basic Book.

Luria, A. R. (1973). *The working brain*. New York, NY: Basic Book.

Maniadakis, M., Trahanias, P., & Tani, J. (2012). Self-organizing high-order cognitive functions in artificial agents: Implications for possible prefrontal cortex mechanisms. *Neural Networks: The Official Journal of the International Neural Network Society*, 33, 76–87. doi:10.1016/j.neunet.2012.04.005

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences*, 15(10), 453–459. doi:10.1016/j.tics.2011.07.011

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (pp. 1–18). New York, NY: Plenum Press.

Podsakoff, P. M., MacKenzie, S. M., Podsakoff, N. P., & Lee, J. S. (2003). The mismeasure of man(agement) and its implications for leadership research. *Leadership Quarterly*, 14, 615–656.

Schoemaker, K., Bunte, T., Wiebe, S. A., Espy, K. A., Dekovic, M., & Matthys, W. (2012). Executive function deficits in preschool children with ADHD and DBD. *Journal of Child Psychology and Psychiatry*, 53(2), 111–119. doi:10.1111/j.1469-7610.2011.02468.x

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 298(1089), 199–209.

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, 44(2), 575–587.

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, 108(3), 436–452. doi:10.1016/j.jecp.2010.08.008

Willoughby, M., Holochwost, S. J., Blanton, Z. E., & Blair, C. B. (2014). Executive functions: Formative versus reflective measurement. *Measurement: Interdisciplinary Research and Perspectives*, 12, 69–95.

Willoughby, M. T., Pek, J., & Blair, C. B. (2013). Measuring executive function in early childhood: A focus on maximal reliability and the derivation of short forms. *Psychological Assessment*, 25(2), 664–670. doi:10.1037/a0031747

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2011). Contributions of modern measurement theory to measuring executive function in early childhood: An empirical demonstration. *Journal of Experimental Child Psychology*, 108(3), 414–435.

Copyright of Measurement is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.